Using Fractal Downscaling of Satellite Precipitation Products for Hydrometeorological Applications

KUN TAO AND ANA P. BARROS

Pratt School of Engineering, Duke University, Durham, North Carolina

(Manuscript received 8 September 2008, in final form 17 September 2009)

ABSTRACT

The objective of spatial downscaling strategies is to increase the information content of coarse datasets at smaller scales. In the case of quantitative precipitation estimation (QPE) for hydrological applications, the goal is to close the scale gap between the spatial resolution of coarse datasets (e.g., gridded satellite precipitation products at resolution \( L \times L \)) and the high resolution \((l \times l; L \gg l)\) necessary to capture the spatial features that determine spatial variability of water flows and water stores in the landscape. In essence, the downscaling process consists of weaving subgrid-scale heterogeneity over a desired range of wavelengths in the original field. The defining question is, which properties, statistical and otherwise, of the target field (the known observable at the desired spatial resolution) should be matched, with the caveat that downscaling methods be as general as possible and therefore ideally without case-specific constraints and/or calibration requirements? Here, the attention is focused on two simple fractal downscaling methods using iterated functions systems (IFS) and fractal Brownian surfaces (FBS) that meet this requirement. The two methods were applied to disaggregate spatially 27 summertime convective storms in the central United States during 2007 at three consecutive times (1800, 2100, and 0000 UTC, thus 81 fields overall) from the Tropical Rainfall Measuring Mission (TRMM) version 6 (V6) 3B42 precipitation product \((\approx 25\text{-km grid spacing})\) to the same resolution as the NCEP stage IV products \((\approx 4\text{-km grid spacing})\). Results from bilinear interpolation are used as the control. A fundamental distinction between IFS and FBS is that the latter implies a distribution of downscaled fields and thus an ensemble solution, whereas the former provides a single solution. The downscaling effectiveness is assessed using fractal measures (the spectral exponent \(b\), fractal dimension \(D\), Hurst coefficient \(H\), and roughness amplitude \(R\)) and traditional operational scores statistics scores \([\text{false alarm rate (FR)}, \text{probability of detection (PD)}, \text{threat score (TS)}, \text{and Heidke skill score (HSS)}\]) as well as bias and the root-mean-square error (RMSE). The results show that both IFS and FBS fractal interpolation perform well with regard to operational skill scores, and they meet the additional requirement of generating structurally consistent fields. Furthermore, confidence intervals can be directly generated from the FBS ensemble. The results were used to diagnose errors relevant for hydrometeorological applications, in particular a spatial displacement with characteristic length of at least 50 km \((2500\text{ km}^2)\) in the location of peak rainfall intensities for the cases studied.

1. Introduction

The case to address the disparity between the spatial and temporal resolution of existing and anticipated satellite-based quantitative precipitation estimates (QPE) and the need for high spatial resolution in hydrometeorological and hydrological applications is amply discussed in the literature (e.g., Bindlish and Barros 2000; Harris et al. 2001; Bell and Kundu 2003; Deidda 2000; Ferraris et al. 2003; among others). Besides spatial sampling, uncertainty in satellite precipitation products also originates from temporal sampling, sensor fidelity, and the retrieval model. In this work, we focus solely on spatial sampling: in particular, we are interested in downscaling methods that explicitly consider the scaling behavior of the spatial statistics of rainfall fields after retrieval.

Although there are many important problems to address in satellite-based measurement of precipitation, including light rainfall, cold-season precipitation, and precipitation at high latitudes in general, we focus here on summertime convective rainfall, which is characterized by high space–time variability and thus poses an additional challenge because of limited temporal sampling (i.e., satellite revisit times). In addition, the operational skill scores...
for areal QPE and quantitative precipitation forecasting (QPF) for this type of rainfall remain very low and have yet to beat persistence at the time scales and for rainfall amounts that matter for hazard warning in practice (e.g., subdaily and 0.5 mm h\(^{-1}\)) as per historical analysis by the Hydrometeorological Prediction Center (HPC; available online at http://www.hpc.ncep.noaa.gov). Much progress has been attained recently via data assimilation of rainfall into operational models; however, even when a dense network of near-real-time observations (rain gauges) is available and can be merged with model simulations such as in the continental United States, skill scores are dramatically lower when model results are considered alone (Davis et al. 2004), especially for warm-season rainfall. In regions where ground-based observations are sparse or are altogether lacking, satellite-based rainfall products provide the only observational basis. It is therefore important to establish pathways to improve their utility.

In this context, the challenge is to define the spatial scale at which rainfall estimates must be produced with accuracy necessary to yield useful hydrological estimates. The impact of the spatial variability of rainfall on hydrologic response, specifically runoff, is illustrated in Fig. 1 for a hypothetical watershed using the area-histogram method. Key elements in flood forecasting are the so-called lag time or time to peak, which is the time since the peak of rainfall until maximum discharge is reached at the watershed outlet; the peak magnitude; the duration of the hydrograph from beginning to end (e.g., the time base); and the overall storm volume (Fig. 1, top right). For example, from the point of view of hazard warning, the lag time is equivalent to the warning lead time, and peak magnitude tests the hydraulic transport capacity of the river channel. Hydrographs (Q1–Q4) were generated for the four rainfall scenarios in Fig. 1, bottom left. Although Q2, Q3, and Q4 show the peak discharge at the same time, albeit with different magnitudes, note how Q1 is completely different from the other hydrographs because of a shift in rainfall from one grid cell to an adjacent one. This example illustrates the importance of getting the precipitation right, where it falls, to be useful for hazard warning and emergency management. Therefore, it turns out that, besides concerns with spatial resolution proper and rainfall amount, another concern in reducing the impact of uncertainty of satellite-based rainfall estimation is that of capturing the correct spatial structure of rainfall. That is the motivation for this work.

The downscaling process provides the means to weave in the original field (scale \(L\)) subgrid-scale heterogeneity (scale \(l\)) over a desired range of wavelengths \([l, L]\). The multiscaling behavior of rainfall has been studied by many (Waymire et al. 1984; Lovejoy and Mandelbrot 1985; Menadbe et al. 1997; Over and Gupta 1996; Deidda 2000; Harris et al. 2001; Ferraris et al. 2003; and many others). In the spectral framework, the downscaling process should lead to a power spectrum of a precipitation field that smoothly merges at the lower wavenumbers with the power spectrum of the observed large-scale field, and it should reproduce at high wavenumbers the power spectra of rainfall fields measured at high spatial resolution. If the scaling behavior between the original field at coarse scale and the desired fine-resolution field is self-affine, then fractal interpolation provides a pathway to downscaling that fits the target objectives (Bindlish and Barros 1996). Deidda (2000) proposed a space–time multifractal method relying on stochastic cascades to downscale sequences of radar fields assuming isotropy and self-similarity in both space and time. Results were evaluated using cumulative density functions (CDFs) of rainfall intensity for different sequences with mixed results depending on storm case, whereas bulk statistical parameters were more consistently reasonable. Highly dynamic variation of fractal parameters during a specific storm was reported by Zepeda-Arce et al. (2000) for a multi-squall-line storm over central Oklahoma and by Bindlish and Barros (2000) for two very different storms in the mid-Atlantic region, over the Susquehanna River basin. Because of the need to determine multifractal parameters a priori, storm-specific and time-varying calibration of space–time scaling parameters, the systematic and generalized implementation of multifractal downscaling faces many challenges. Rebora et al. (2006) presented a space–time algorithm very similar to Bindlish and Barros (1996), including the time dimension. The method was applied to generate a distribution of downscaled rainfall fields from model forecasts of a specific storm. The downscaled rainfall fields were subsequently used to force a hydrology model to generate a predictive distribution of flood peak discharge at a location where hydrometric observations were available; however, detailed evaluation and/or verification of the rainfall fields produced by their algorithm against observations is lacking (e.g., Rebora et al. 2006; Siccardi et al. 2005; Ferraris et al. 2002). Generally, the advantage of fractal downscaling methods in their different implementations is that, by preserving scaling behavior, higher-order space–time statistics (moments of the distribution) and therefore spatial and temporal structure are preserved.

Here, we explore the performance of two alternative fractal interpolation methods: iterated functions system (IFS) and fractal Brownian surfaces (FBS). Although these methods can be modified in multiple ways by either calibration or postprocessing to improve performance for specific locations and/or specific storm
types or for specific scales (e.g., Bindlish and Barros 1996, 2000; Kuligowski and Barros 2001; Kim and Barros 2002), here we rely on the simplest formulation that stems strictly from mathematical statistics to eliminate any effect of implicit tuning or heuristic adjustment in the application of the methods; that is, the fractal downscaling algorithms explored here are strictly data driven, mathematically simple, and general and there is no parameter, structural function, or stochastic synthetic rainfall-intensity model that needs to be calibrated. The premise is that the downscaling method should not be a source of ambiguity in the characterization of uncertainty at different scales and thus should be a neutral player in error propagation analysis to assess the utility of satellite precipitation products for hydrological applications.

The two methods were applied to disaggregate spatially 27 summertime convective storms in the central United States during 2007 at 1800, 2100, and 0000 UTC (thus 81 fields overall) from the Tropical Rainfall Measuring Mission (TRMM) version 6 (V6) 3B42 precipitation product (~25 km grid spacing) to the same resolution as the National Centers for Environmental Prediction (NCEP) stage IV products (~4 km grid spacing). Results from bilinear interpolation are used as the control reference, in that the small range of scales (~1/5 ratio) and the nature of the 3B42 data (includes information from the same rain gauges as the NCEP stage IV product) should provide an optimal setup. The downscaling performance is assessed using fractal metrics: the spectral exponent β that captures the wavenumber distribution; the fractal dimension D to describe the space-filling

FIG. 1. Rainfall-runoff response (top left) using area-histogram method to derive hydrographs (streamflow Q) generated from a stylized watershed (bottom left) under a hypothetical rainfall forcing of 200 units of rainfall distributed according to four different scenarios. Thick black lines on the watershed indicate the stream network. The watershed is assumed impermeable (no infiltration or other losses), and rainfall is uniform over each grid cell (shaded squares). The arrows indicate runoff direction over land, and the flow in the streams (thick black lines) is toward the outlet (triangle). Fractions represent the portion of the grid cell that reaches each of the immediately adjacent streams, with each corresponding to a side of a grid cell and one time unit in terms of flow propagation. Fractions indicate the portion of the pixel that flows in the arrow direction. Italic node numbers indicate the distance away from the basin outlet (blue triangle) in units of time Δt. Note the differences resulting from variation in rainfall intensity but also in spatial distribution with regard to the stream network. This figure illustrates the impact of potential systematic spatial distribution errors on the estimation of the peak and magnitude of streamflow (and flood) response.
characteristics; the Hurst coefficient \( H \) to quantify long-range spatial correlation; and the roughness amplitude \( R \) that is a measure of spatial variance at the nominal resolution. Traditional operational scores statistics scores are also used, including the false alarm rate (FR), the probability of detection (PD), the threat score (TS), and the Heidke skill score (HSS), as well as bias and the root-mean-square error (RMSE). The purpose is not to show that the methods work or are superior alternatives to bilinear interpolation of TRMM 3B42 in the United States, but rather to establish a robust framework for downscaling in regions of the world where ground observations (e.g., rain gauges and radars) are lacking.

The article is organized as follows: a description of the data is provided in section 2, and the methodology is introduced in section 3, with particular emphasis on describing IFS. The analysis of the results is given in section 4, and section 5 provides a summary and conclusions.

2. Data

Datasets used in this work are the TRMM 3B42 V6 (available online at http://trmm.gsfc.nasa.gov/3b42.html) precipitation product, a 3-hourly dataset at 0.25° × 0.25° resolution obtained by merging multisensor calibrated passive microwave data and infrared precipitation estimates, which are combined and subsequently rescaled to match monthly rain gauge statistics (Huffman et al. 2007), and the NCEP stage IV analysis (available online at http://www.emc.ncep.noaa.gov/mmb/ylin/pcpanl/stage4/). The NCEP stage IV is a national rainfall product mosaicked at NCEP and derived from the regional hourly/6-hourly multisensor (radar + gauges) precipitation analyses (MPEs) produced by the 12 river forecast centers over the conterminous United States (CONUS). A total of 27 convective storms that produced heavy rainfall in south-central Oklahoma (29.125°–38.875°N, 100.875–88.125°W) between May and September of 2007 were used for this study (Table 1). Each storm was followed over a 9-h period, which started with 3-hourly accumulations at 1800 UTC (1200 LST) and ended 6 h later at 0000 UTC (1800 LST) in order to span the entire duration of the storm over the region of study. The average of the three NCEP stage IV hourly fields corresponding to each 3-hourly TRMM 3B32 field are used as the truth; a total of 81 of the 3-hourly rainfall fields were downscaled.

3. Methodology

In this work, we aim at deriving precipitation fields at high spatial resolution by fully exploring the fractal nature of the original 3B42 field. For self-affine processes, the spectral exponent \( \beta \), fractal dimension \( D \), and Hurst coefficient \( H \) are three interchangeable characteristics; the latter two can be easily derived from the first parameter as follows for a 3D field \((x, y, \text{rainrate})\):

\[
D = \frac{7 - \beta}{2} \quad \text{and} \quad (1)
\]

\[
H = \frac{\beta - 1}{2}.
\]

The spectral exponent, which is defined as the slope of the log–log plot of a 2D spatial Fourier power spectrum, describes the variability of the rainfall fields as a function of wavelength and/or spatial scale, whereas the roughness amplitude, the power at the shortest wavelength, measures the variance of the rainfall field at the corresponding spatial scale. The Hurst coefficient, which is also referred to as the self-affinity index, measures long-range correlation: processes with long-range spatial persistence correspond to \( 1 \geq H > 0.5 \); for anti-persistent processes, \( 0 \leq H \leq 0.5 \); for white noise, \( H = 0 \); and \( H < 0 \) for fractional noise (highly singular processes without characteristic scales). Therefore, assuming that rainfall fields are self-affine processes by matching the

<table>
<thead>
<tr>
<th>Storm</th>
<th>Month</th>
<th>Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>May</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>May</td>
<td>2</td>
</tr>
<tr>
<td>3*</td>
<td>May</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>May</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>May</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>June</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>June</td>
<td>14</td>
</tr>
<tr>
<td>8</td>
<td>June</td>
<td>15</td>
</tr>
<tr>
<td>9</td>
<td>June</td>
<td>25</td>
</tr>
<tr>
<td>10</td>
<td>June</td>
<td>26</td>
</tr>
<tr>
<td>11</td>
<td>June</td>
<td>27</td>
</tr>
<tr>
<td>12</td>
<td>June</td>
<td>28</td>
</tr>
<tr>
<td>13</td>
<td>June</td>
<td>29</td>
</tr>
<tr>
<td>14</td>
<td>July</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>July</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>July</td>
<td>3</td>
</tr>
<tr>
<td>17</td>
<td>July</td>
<td>6</td>
</tr>
<tr>
<td>18*</td>
<td>July</td>
<td>9</td>
</tr>
<tr>
<td>19</td>
<td>July</td>
<td>12</td>
</tr>
<tr>
<td>20</td>
<td>July</td>
<td>23</td>
</tr>
<tr>
<td>21</td>
<td>July</td>
<td>28</td>
</tr>
<tr>
<td>22</td>
<td>July</td>
<td>30</td>
</tr>
<tr>
<td>23</td>
<td>August</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>August</td>
<td>17</td>
</tr>
<tr>
<td>25</td>
<td>August</td>
<td>24</td>
</tr>
<tr>
<td>26</td>
<td>August</td>
<td>8</td>
</tr>
<tr>
<td>27*</td>
<td>August</td>
<td>26</td>
</tr>
</tbody>
</table>

* Storm used as an example in this paper.
spectral slope and the roughness factor of the referenced 3B42 rainfall fields, a modified FBS \((H = 0.5)\) can be used as the interpolating surface to extend the power spectrum to shorter wavelengths:

\[
Z(u, v) = \frac{Z_0(u, v)}{k_r^{(\beta-\beta_y)^2}} \exp \left\{ \frac{1}{2} \left[ S_{r,1} - \frac{\beta + 1}{\beta_b + 1} S_{r,2} \right] \right\},
\]

where \(Z(u, v), Z_0(u, v)\), \(\beta\), and \(\beta_b\) are the Fourier transforms of the interpolating and the Brownian surfaces, as well as the referenced and the Brownian power spectral exponents, respectively; \(k_r\) is the radial wavenumber; and \(S_{r,1}\) and \(S_{r,2}\) denote the roughness factors of the original and the Brownian noise surface. Detailed descriptions of different implementations of this approach can be found for different applications in Bindlish and Barros (1996, 2000, 2002). As noted by Kim and Barros (2002), the construction of fractal interpolation surfaces using FBS has two characteristics: 1) it can only be applied within a range of spatial scales where the data are self-affine (single scaling), and 2) the downscaled surfaces are not unique solutions, merely one of the many possible realizations of a random process. Therefore, a distribution of surfaces must be generated to determine the ensemble-mean interpolation surface. Alternatively, an ensemble of downscaled fields can be created using each interpolation surface separately as done here and using the mean of the ensemble for evaluation purposes. Bindlish and Barros (2002), Kuligowski and Barros (2001), and Barros and Tao (2008) used observed fields of proxy variables at higher spatial resolution as interpolating surfaces, which takes care of the need to produce an FBS ensemble; however, this is not possible when operating in measurement or prognostic mode without the support of ground observations. This is important, because there are no ground observations of rainfall or proxy alternatives (i.e., ground truth) over many regions of the world to calibrate downscaling algorithms, as opposed to what happens in the United States.

An alternative to the traditional FBS interpolation was proposed by Barnsley (1986) using IFS as fractal interpolation functions of self-affine data. The IFS are deterministic and are defined uniquely by the existing coarse resolution data. Xie and Sun (1997) adopted the concept to contraction mapping by proposing a bivariate IFS of which the roughness and fractal dimension can be adjusted to match the behavior of observations using a vertical scaling factor. Kim and Barros (2002) further modified the method to incorporate spatial non-stationarity by repeatedly applying the method on a sliding window rather than globally to an entire field and to include spatially and temporally varying vertical scaling factors. Here, however, no adjustable vertical scaling factor was used, because no calibration is conducted. A detailed mathematical description of the method is provided by Xie and Sun (1997) and Kim and Barros (2002). The algebraic form of the IFS used in this work is provided later.

Consider a 3D field of data points \((x, y, z)\) corresponding to an \((M \times N)\) matrix. The new interpolated points are generated using affine transformations defined by the functions \(W\) (see Fig. 2 for illustration):

\[
\begin{align*}
W_{mn}(x) &= a_{mn}x + b_{mn}, \\
W_{mn}(y) &= c_{mn}y + d_{mn}, \\
W_{mn}(z) &= e_{mn}z + f_{mn}y + g_{mn}xy + h_{mn},
\end{align*}
\]

where \(m = 1, 2, \ldots, M, n = 1, 2, \ldots, N; a_{mn}, b_{mn}, c_{mn}, d_{mn}, e_{mn}, f_{mn}, g_{mn}, h_{mn}\) are calculated as follows:

\[
\begin{align*}
a_{mn} &= \frac{x_m - x_{m-1}}{x_M - x_0}, \\
b_{mn} &= \frac{x_0 - x_M}{x_M - x_0}, \\
c_{mn} &= \frac{y_n - y_{n-1}}{y_N - y_0}, \\
d_{mn} &= \frac{y_0 - y_N}{y_N - y_0}, \\
e_{mn} &= \frac{(z_{m,n-1} - z_{m-1,n-1})y_{n-1} - (z_{m,n} - z_{m-1,n})y_0}{(x_M - x_0)(y_N - y_0)}, \\
f_{mn} &= \frac{(z_{m,n-1} - z_{m-1,n-1})x_{n-1} - (z_{m,n} - z_{m-1,n})x_0}{(x_M - x_0)(y_N - y_0)}, \\
g_{mn} &= \frac{z_{m,n-1} - z_{m,n-1} - z_{m-1,n} + z_{m-1,n-1}}{x_M - x_0}(y_N - y_0), \\
h_{mn} &= \frac{z_{m,n}x_0y_0 - z_{m,n-1}x_Ny_0 - z_{m-1,n}x_Ny_0 + z_{m-1,n-1}x_Ny_N}{(x_M - x_0)(y_N - y_0)}.
\end{align*}
\]

In our case, each 3D point \((x, y, z)\) represents latitude, longitude, and 3-hourly rainfall accumulation. In Fig. 2a, shaded circles denote the known points of the original field (coarse resolution) and filled circles represent the location of the interpolated points. Using the affine transformations defined by Eq. (3), the scaling behavior within the subregion \((x_{m-1}, y_{n-1}), (x_{m-1}, y_n), (x_m, y_{n-1}), (x_m, y_n)\), and \((x_m, y_0)\) is preserved and thus remains self-affine with its parent field defined by \((x_0, y_0), (x_M, y_0), (x_0, y_N), \) and \((x_M, y_N)\).

In practice, it is important to determine the proper range of scales (downscaling window) over which the property of self-affinity holds. In other words, the size of the parent field affects the spatial look of the interpolated results to some degree. In terms of the 3B42 rainfall data, generally, the larger the parent field, the fewer visually identifiable spatial structures. Various
downscaling window sizes were experimented with, both globally and locally, and eventually a $3 \times 3$ mapping window was selected; that is, we assume that the precipitation pattern within a $75 \times 75$ km$^2$ area in the original 3B42 is self-affine. The interpolation is recursive, first from the original resolution 25 km to the intermediate resolution 8.3 km and then from 8.3 km to the final 4.2 km. Finally, bilinear interpolation was carried out using supporting functions of the form

$$f(x, y) = B_1 + B_2x + B_3y + B_4xy.$$  \hspace{1cm} (4)

Along the external boundaries (dotted circles in Fig. 2b), the average of the two adjacent values was used to
FIG. 3. 3B42 V6, NCEP stage IV, IFS fractal downscaled 3B42 V6, Brownian fractal downscaled 3B42 V6, and linear downscaled 3B42 V6 precipitation fields at 0000 UTC 26 Sep 2007 (storm event 27, Table 1).
FIG. 4. Log-log power spectrum density function (PSD) plots of 3B42 V6, NCEP stage IV, and downscaled precipitation fields with various algorithms at 0000 UTC 26 Sep 2007. The radial wavenumber is $\sqrt{u^2 + v^2}$, where $(u, v)$ are spatial indices that are $(0, 0)$ at the center of the rainfall field over the region of interest. The FBS result corresponds to the ensemble mean of 100 replicates. The 95% confidence intervals for $D$ and $H (H = Ha)$ are $\pm 0.01$, $\pm 0.02$ for $\beta$, and $\pm 0.03$ for $R$. 
calculate the downscaled rainrate at \((X_b, Y_b)\); otherwise, calculations were made first for centroids \((X_c, Y_c)\) of the coarse-scale grid elements (filled circles in Fig. 2b),

\[
B_1 = f(X_{m-1}Y_{n-1}),
\]

\[
B_2 = f(X_mY_{n-1}) - B_1,
\]

\[
B_3 = f(X_{m-1}Y_n) - B_1, \quad \text{and}
\]

\[
B_4 = B_1 - f(X_mY_{n-1}) - f(X_{m-1}Y_n) + f(X_mY_n),
\]

and second for the intermediate points \((X_i, Y_i)\) in the internal boundaries (dashed circles in Fig. 2c),

\[
B_1 = f(X_cY_c),
\]

\[
B_2 = f(X_mY_{n-1}) - B_1,
\]

\[
B_3 = f(X_mY_n) - B_1, \quad \text{and}
\]

\[
B_4 = B_1 - f(X_mY_{n-1}) - f(X_{m-1}Y_n) + f(X_cY_{c+1}).
\]

Areas where no rain was present in the original 3B42 fields were masked out as rain-free areas for all downscaling methods. When the rain-free mask was not applied, the performance of IFS and linear interpolation was very poor because of the generation of extensive regions with artificial very light rainfall.

4. Analysis of results

As mentioned in section 2, there are 27 storm days with three 3-hourly rain accumulation fields per day and thus 243 downscaled fields with 81 for each interpolation method. Figure 3 shows one example of such fields including the original 3B42 V6; the corresponding NCEP

---

**Fig. 5.** Intercomparison of fractal parameter statistics for all 81 summertime afternoon convection fields: (a) \(\beta\); (b) \(R\); (c) \(D\); and (d) \(H\).
stage IV field; and the downscaled IFS, FBS, and linear interpolation fields at 0000 UTC 26 September 2007. The FBS rainfall field is the ensemble mean of a distribution of 100 downscaled FBS rainfall fields. A cursory visual survey can capture the apparent difference between the smoother fields produced by bilinear interpolation as compared to the fractal interpolation methods. The log–log plots of the power spectra of the corresponding datasets are shown in Fig. 4. Compared with the original 3B42, power spectral exponents from various downscaled fields are closer to those estimated from the NCEP data. Note that the 95% confidence intervals (±0.01 for $D$ and $H$, ±0.02 for $\beta$, and ±0.03 for $R$) are very small and cannot be clearly distinguished in the plots. The values were of similar order of magnitude for all fields. This case was randomly selected, and it does reflect the overall performance of any particular method. For the remainder of the manuscript all calculations including skill scores and residual estimates are presented for the ensemble mean only. The scatterplot of rainfall between NCEP and various downscaled precipitation fields pixel by pixel (not shown) is similar to scatterplots shown by Barros and Tao (2008) for other case studies, corresponding to a cloud of points spreading out on both quadrants below and above the 1:1 line indicating the presence of regions where the downscaled rainfall and TRMM 3B42 are in strong disagreement: one is high and the others are low and vice versa. This issue
will be revisited later when the morphology of spatial residuals is analyzed.

A summary of performance for all 81 cases is presented in Figs. 5a–d by means of the scatterplots of four fractal parameters: $b$, $D$, $H$, and $R$. Note the spread in the comparison of the fractal statistics of the original TRMM 3B42 V6 fields against the NCEP stage IV, whereas the results are tightly clustered for both fractal interpolation methods and more so for the IFS than the FBS. This reflects not only the robustness of the downscaling methodology but also the consistency in the generation of the TRMM 3B42 fields (Huffman et al. 2007); because the information of spatial variability near $0.25^\circ \times 0.25^\circ$ is consistent with reality, the IFS and FBS can capture it so well. The bilinear interpolation performs inconsistently from storm to storm with large spread in the fractal

Fig. 7. Demonstration of scale effects associated with aggregation of the NCEP stage IV field. The dominance of the light rainfall effect over the skill scores is illustrated at 2100 UTC for storm event 27. Also, note that detection is a problem at 1800 UTC (1200 LST) and significantly improves later in the afternoon.
parameters, depending on storm characteristics. The results show that bilinear interpolation arbitrarily adds spatial persistence in the downscaling process, as illustrated by the large spread of the Hurst coefficient. Note that the Hurst coefficient for the NCEP fields is largely at or above 0.5 (Fig. 5d). By contrast, the original TRMM 3B42 exhibits fractional noise behavior with $H < 0$ for a number of early afternoon fields with scattered convective activity at only a few single isolated pixels very far apart from each other, because the characteristic spatial scale of convective cells is smaller than the spatial resolution of the dataset. Consider the anomalous behavior in Fig. 5c, which shows values of fractal dimension $D > 3$ in the case of the TRMM 3B42. This is an artifact of the approach used to estimate the fractal dimension [implicit in Eq. (1)]: $D + H = 3$, which is only strictly valid for $0 \leq H \leq 1$ (Hansen et al. 2001). Generally, the fractal dimension and the Hurst coefficient should be calculated independently of each other (Gneiting and Schlather 2004). On the other hand, unexpected behavior such as that identified here for the TRMM 3B42 can be used as a quality-control diagnostic to identify those conditions under which the retrieval algorithm and subsequent postprocessing are particularly challenging.

Inspection of Figs. 5a–d shows that fractal metrics for both IFS and FBS are tightly clustered (more so for FBS than IFS) close to the 1:1 line with respect to the NCEP stage IV data, although there is evidence of some bias. Recall that, using either fractal downscaling method, the algorithm relies alone on the scaling behavior that can be inferred from the spectral slope of the 3B42 dataset over a limited range of scales (roughly 100–25 km); therefore, information content that is not present in the

---

**Fig. 8.** Categorical skill scores for all downscaled storm events in Table 1 as a function of time of day and rainfall-intensity threshold for the 3-h period centered at the times indicated: (a) bias; (b) HSS; and (c) TS [TS = critical success index (CSI) = 0.5 indicates 50% of success]. Values between ( ) indicate average skill scores for the corresponding time.

**Fig. 9.** RMSE of rainfall intensity for each downscaled storm event for each 3-h period centered at 1800, 2100, and 0000 UTC.
TRMM 3B42 field itself cannot be artificially generated, because there is no calibration or postprocessing. That is, ultimately the quality of the downscaled fields depends on the intrinsic quality of the original dataset at coarse resolution: the better the estimates of precipitation at coarse resolution, the better the downscaled products will be. In this context, the results are indeed good by comparison against the distribution of fractal parameters for bilinear interpolation, for example, which exhibits smoother gradients and artificially spread areas of light rainfall, thus amplifying spatial persistence (high $H$ values) and decreasing variance (low $R$ values).
Although the fractal parameters provide insight with regard to spatial structure, standard operational statistical scores provide a measure of the bulk effectiveness of the downscaling algorithm. Given a certain rainfall threshold (0.25 and/or 0.5 mm hr$^{-1}$) and using a two-way contingency table of satellite-based rainfall estimates versus the NCEP stage IV gridded observations, the data are organized into four classes: an instance when both satellite product and NCEP stage IV grid points match or exceed a specified rain rate threshold is a hit (H); when the opposite happens, the estimate is classified as a miss (M) if the NCEP value matches the criterion and a false alarm (FA) otherwise. The statistical measures are obtained by combination of these classes as follows:

(i) false alarm rate: \( FR = FA/(H + FA) \), \( 0 \leq FR \leq 1 \);
(ii) probability of detection: \( PD = H/(H + M) \), \( 0 \leq PD \leq 1 \);
(iii) threat score: \( TS = H/(H + FA + M) \), \( 0 \leq TS \leq 1 \);
(iv) and the Heidke skill score \( [ -1 \leq HSS \leq 1 ] \):

\[
HSS = \frac{2 \times (Z \times H - FA \times M)}{[(H + FA) \times (Z + FA)] + [(M + H) \times (M + Z)]},
\]

where \( Z \) is the overall number of zeros (when neither product nor observations match the specified threshold criterion). In addition, rain accumulation bias and RMSE are also calculated. A Heidke skill score of zero implies no better skill than random chance, whereas a TS of 0.5 as defined here indicates that the assessment criterion is met at least 50% of the time. These statistical measures should be viewed as indicators of relative performance and not absolute quantitative descriptions of data fitness. Further details and insights on the use of these
statistical measures, please see the postings of the National Precipitation Verification Unit (NPVU; available online at http://www.hpc.ncep.noaa.gov/npvu) or the Web site maintained by the World Weather Research Program Joint Working Group on Verification (Ebert et al. 2005; McBride and Ebert 2000; Ebert and McBride 2000). Casati et al. (2008) provide a comprehensive review of the challenges in the evaluation of QPF, which are in many ways similar to those faced in QPE. Here, in order to identify spatial scale effects, the skill scores were calculated for all downscaled fields versus the NCEP stage IV fields at their original resolution and for the original TRMM 3B42 V6 versus the NCEP stage IV aggregated to the same resolution by simple averaging in a sensor-like fashion. Generally, there was no substantial difference among different downscaling methods in the skill scores except for the lower rainfall threshold, which reflects the measurement challenge of light rainfall detection and isolated showers from satellite-based sensors.

The scatterplots of probability of detection versus false alarm rates provide a visual measure of discrimination skill or product utility; useful measurements are above the 1:1 line (Figs. 6a,b). The ideal behavior coincides with clustering at the top-left corner of the scatterplot. In Fig. 6a, the data correspond to the downscaled TRMM 3B42 fields vis-à-vis the nominal NCEP stage IV; in Fig. 6b, the data correspond to the original TRMM 3B42 resolution and the aggregated NCEP stage IV fields. The improvement in reliability for the downscaled fields is significant for the rainfall detection threshold (0 mm hr$^{-1}$), because light rainfall is largely missed for many storms by TRMM 3B42 V6, even when the ground observations indicate that it is widespread, as shown in the example in Fig. 7a. Figure 7b illustrates the effect of artificial spreading of light rainfall (<0.5 mm hr$^{-1}$), especially for IFS and linear interpolation. Overall, the plots highlight the value of the satellite observations with a systematic improvement from 1800 to 0000 UTC (from 1200 to 1800 LST), because 1200 LST rainfall is rare in summertime and when it occurs it is characterized by scattered showers and low rain rates.

The results at 1800, 2100, and 0000 UTC (1200, 1500, and 1800 LST) for selected skill scores are shown in Fig. 8. It is important to remind the reader at this time that the TRMM 3B42 product includes bias corrections at the monthly time scale based on the same rain gauges that contribute to the NCEP stage IV analysis. Therefore, assuming that each grid cell is one independent sample, as is the case with categorical scores, the baseline for each measure of skill is strongly determined by the quality of the 3B42 original field and differences among downscaling methods should be small, especially in this case when the scale ratio is $\sim$1:5. Indeed, there is no significant difference among the three downscaling approaches in terms of categorical skill scores. One encouraging result is that the skill scores do change for higher rainfall thresholds (e.g., ΔTS for 0.5 mm hr$^{-1}$ threshold versus 0.25 mm hr$^{-1}$ is on average 0.1), but for each threshold the performance is consistent at all different times (1200, 1500, and 1800 LST), with exceptions (±) for a handful of cases, such as the examples discussed later to illustrate the different morphologies of error. Finally, note how, despite the fact that the categorical indices were calculated at the storm scale and therefore no averaging is involved, they compare favorably with the performance documented in the literature for satellite-based data (e.g., Huffman et al. 2007; Joyce et al. 2004; Kidd et al. 2003; among others) and the skill scores published by HPC.

As shown in Fig. 9, the RMSE is somewhat larger for FBS for all cases, especially for 2100 UTC fields. This
begs the question of whether this is a result of using a very large area, including a large number of no-rain pixels, to determine these skill scores, which might vary significantly in space depending on storm characteristics if the calculations were limited to the most intense storm areas. Interestingly, there is little difference between IFS and bilinear interpolation, which is also consistent with the histograms in Fig. 7b. The one aspect that IFS and bilinear interpolation have in common is that both weigh heavily local information, where FBS takes a holistic approach of the scaling behavior of the fields over the entire region of study.

Residual fields at 0000 UTC 26 September 2007 are shown in Fig. 10a to illustrate the discrepancy between NCEP stage IV and TRMM 3B42 estimates independently of downscaling method. One important feature of these residual maps is to show the dipole of large positive/negative residuals oriented along the main axis of the squall line centered at (34°N, 96°W) (Fig. 3) corresponding to a characteristic displacement length of about 50 km. Although with lesser magnitude, similar behavior can be found in the vicinity of all heavy precipitation areas in Fig. 3 (top right). This suggests a systematic zonal placement error in the original 3B42 field that can probably be investigated by unscrambling the specific combination of satellite sensors that contribute the rainfall estimates for each field to identify the source of the systematic spatial shift, which might be related to overpass geometry and sensor-specific algorithms. Note how this type of critical geolocation error would not be detected by global analysis of the empirical probability distributions for all fields (Fig. 10b). Nevertheless, this spatial shift in the location of peak rainfall is the type of error that would cause dramatic differences in rainfall-runoff response as illustrated in Fig. 1 for basins smaller than 2500 km². Otherwise, in the case of IFS, a major problem is that of handling small disconnected storms (thunderstorms and isolated showers), because there is no spatial connectivity to distribute the excessive rainfall. In such cases, if the TRMM 3B42 V6 overestimates or underestimates rainfall accumulations, the downscaling method does not improve substantially on the original fields. On the other hand, it is critical that the downscaling methodology does not add to or compound the spatial errors, and this provides the impetus to use downscaling techniques such as IFS that preserve spatial structure systematically.

Categorical skill scores such as those presented here have often been criticized by their lack of ability to assess other storm attributes such as orientation, coverage area, and variance (Ebert and McBride 2000; Göber et al. 2003; Schaefer 1990; Davis et al. 2006). On the other hand, they do provide useful measures of overall behavior that are not data or case-study dependent. They also provide a measure, although imperfect, of reliability and confidence (e.g., TS > 0.5). Therefore, they can be viewed as a quality-control flag of sorts to evaluate and improve retrieval algorithms systematically. Take, for example, the very poor HSS for storm event 3 (9 May 2007), suggesting that there is almost no difference between TRMM 3B42 estimates and a roll of the dice (random chance) at 1800 UTC, as compared to the average results at subsequent times. The 3B42 and NCEP rainfall fields at 1800 and 2100 UTC are shown at their original spatial resolutions (0.25° × 0.25° and 4 km × 4 km, respectively) in Fig. 11a, and the corresponding histograms are presented in Fig. 11b. In Fig. 11, right, two example areas where disagreement is large at 2100 UTC are circled in black. These relate to two different manifestations of error: (i) error by excess (areal expansion and morphology) and (ii) error by default (magnitude, location, and morphology). Inspection of the histograms in Fig. 11b (bottom) suggests that IFS and linear interpolation actually compare well with NCEP for very low rainfall rates in this case, which is explained by error compensation: the large area of light rainfall in the observations that is missed by 3B42 is recaptured artificially via the downscaling process. In contrast, note the disparity between the NCEP histogram and the histograms of all other fields in Fig. 11b (top panel). A survey of the left panels in Fig. 11a explains why the scores are so low at 1800 UTC: the combination of sensors used to generate the 3-hourly field completely missed critical storm cells. This is an error by omission (smoothing by default) that cannot be resolved by downscaling, only by improved sensing and retrieval. A different manifestation of overestimation error without changing the overall area affected by the storm is presented in Fig. 12 for storm event 18 (Table 1), which corresponds to the very high RMSE values at 2100 UTC 9 July 2007 (Fig. 9). In the areas of high rainfall, the 3B42 combination product bridges isolated rain cells to form a very large cell, and all the light rainfall areas in between were eliminated in the 3B42 data (smoothing by excess). The areas of light rainfall surrounding the heavy precipitation cores were artificially expanded laterally for all downscaled fields (Fig. 12b).

5. Final remarks

The results show that the both IFS and FBS interpolation outperform the optimal bilinear interpolation with regard to spatial statistics, whereas all three methods have nearly equivalent categorical skill scores, which reinforces the need for caution in the use of categorical
scores alone. It also may explain the conspicuous use of linear interpolation in the analysis of rainfall fields in the verification and intercomparison of quantitative precipitation estimates (QPE) and forecasts (QPF) when high-spatial-resolution data are available.

From the point of view of general applications, especially in regions of the world where ground-based observations are not available, an implementation of IFS downscaling incorporating local stretching and contraction parameters as a function of rainfall intensity similar to Kim and Barros (2002) should be able to capture spatial singularities and strong gradients, although this would require adaptive parameter calibration. The implementation of FBS downscaling is straightforward, because it only depends on simple linear transformations in Fourier space; because it implies the simulation of a distribution, it is directly amenable to generate confidence limits in verification results, thus providing a measure of uncertainty that is highly desirable in quantitative precipitation verification (Casati et al. 2008). Furthermore, the ensemble of high-resolution fields generated by FBS downscaling can in turn be used to force a flood forecasting model to produce probabilistic quantitative flood forecasts (PQFF) within physically meaningful consistency bounds. Although this ensemble approach stops short of achieving the characterization of the total predictive probability distribution function as defined by Krzysztofowicz (2001) in that...
it does not include uncertainty information associated to the retrieval algorithm proper in the case of QPE or a numerical model in the case of QPF. It does provide a measure of the most critical source of uncertainty in hydrological applications. In addition, in the case of small headwater catchments, the ensemble rainfall field can be used directly to produce probabilistic flash flood guidance. Nevertheless, zonal displacement errors were identified that might compromise the utility of satellite products at spatial scales below the characteristic displacement length (~50 km or basins of 2500 km²) for this region. Of course, to establish the generality of this result, similar analysis should be conducted for a much larger number of storms, in different seasons of the year, and for different geographic locations. For example, there may be some systematic types of measurement errors that can be identified and addressed by following a strategy similar to that presented in this study. This is an important result, because hydrometeorological hazards are generally singular events not suitable to spatial averaging: at a specific location, a flood (or a landslide) is a binary event—it either happens or it does not.

Finally, a survey of skill scores provides an opportunity to reflect on the utility of satellite precipitation products vis-à-vis the state of the art in operational QPE. Take, for example, the threat score (TS) of NWP models (e.g., WRF; Gallus 2002; Davis et al. 2004, 2006; Roberts 2008) for summertime convective rainfall that is on average between 0.20 and 0.25 for about 7 mm (0.25 in.) in 24 h. Despite limitations, the TS values of satellite products for the same amount of rainfall are on average well above 0.4, with the additional handicap of being exactly matched within a 3-h window. Based on historical analysis of the rate of improvement of model and postprocessing skill, the difference could amount to a 30-yr lag between satellite and NWP estimates (e.g., HPC). Thus, there is no question of the value of satellite QPE, especially for regions where ground-based observations are scarce.

The development of satellite products of rainfall per se is a complex and challenging endeavor (Hsu et al. 1997; Janowiak et al. 2001; Huffman et al. 2007; and many others). Presently, a degree of maturity has been achieved that should facilitate further improvement of these products in different ways: first, by improving sensor-specific retrieval algorithms as well as multisensor algorithms that recognize and account for the observed spatial structure (e.g., Barros and Tao 2008); second, by improving the spatial resolution of rainfall products without degenerating the spatial information content at the nominal retrieval scale.

Acknowledgments. This work was supported by NASA Grant NNG04GP02G with the second author through the Precipitation Measurement Missions program. We thank the reviewers for thoughtful insights and suggestions. The database of downscaled storms and the downscaling algorithms are available upon request.

REFERENCES


